Clasificador de logs de acceso para detección de incidentes de ciberseguridad

Customized access log classifier for cybersecurity incident detection

Miguel Pérez del Castillo [1], Gastón Rial [2], Rafael Sotelo [3], Máximo Gurméndez [4]

Recibido: Diciembre 2019 Aceptado: Noviembre 2019

Resumen.- Recientemente los sitios web de los gobiernos en el mundo han sido objeto de ataques informáticos. Por ello urge una solución que asista a los analistas de ciberseguridad a detectar los incidentes con rapidez. Para optimizar el tiempo de detección en el proyecto se desarrolló un clasificador que filtre líneas de logs de servidores web en formato CLF (Combined Log Format) que indican comportamiento anómalo. Para ello, se codifican los logs de acceso en representación vectorial y luego se usa el algoritmo de aprendizaje automático K-NN ponderado (K vecinos más próximos) para filtrar los logs. Los datos de entrada fueron provistos por el CERTuy (Equipo de Respuesta ante Emergencias Informáticas) y el SOC (Centro de Operaciones de Seguridad). De las pruebas realizadas sobre el servicio de clasificación, se detectó el 82% de ofensas de ciberseguridad de un conjunto de datos asociado, se logró filtrar el 80% de logs que indican comportamiento normal y se disminuyó el tiempo de detección de logs que indican comportamiento anómalo de 13 horas a 15 minutos.

Palabras clave: Filtrado; Respuesta de ciberseguridad; CLF; Aprendizaje automático.

Summary.- *The number of attacks on government websites has escalated in the last years. In order to assist in the detection process conducted by cybersecurity analysts, this document suggests implementing machine learning techniques over web server access logs. The overall objective is to optimize the detection time using a customized classifier which selects traces corresponding to anomalous activity. Specifically, web server combined log format (CLF) access logs coded as real vectors are an input to a weighted K-NN nearest neighbors' model. The methodology was tested on datasets and premises provided by the CERTuy (National Cybersecurity Event Response Team) and the SOC (Security Operations Center). According to evaluations 82% of cybersecurity offenses have been detected, 80% of normal behavior has been filtered and the reduction time has been reduced from 13 hours to 15 minutes.*

Keywords: *Filtering, Cybersecurity response, CLF, Cachine learning*.

1. Introducción.-

Desde el 2010 la digitalización de los procesos en el ámbito estatal y corporativo, además del incremento acelerado del número de usuarios web, debido a desarrollos tecnológicos en los medios de acceso e infraestructura de telecomunicaciones, ha dejado vulnerables a los portales estatales, los cuales son blanco para el cibercrimen y el espionaje internacional.

[1] Ing. Informática, Universidad de Montevideo, mperez9@correo.um.edu.uy, ORCID iD: 0000-0001-5500-8892

[2] Ing. Telemática, Universidad de Montevideo, grial1@correo.um.edu.uy, ORCID iD: 0000-0001-9174-5937

[3] Dr. Ing. Telemática, Universidad de Montevideo, rsotelo@um.edu.uy, ORCID iD: 0000-0002-4034-3177

[4] Ing. Informática, Universidad de Montevideo, mgurmendez@correo.um.edu.uy, ORCID iD: 0000-0001-6435-0200

El desarrollo y supervisión de esos portales web es impulsado por AGESIC (Agencia para el Gobierno Electrónico y la Sociedad de la Información y el Conocimiento). Entre otras tareas dirige la estrategia e implementación del gobierno electrónico en Uruguay, y controla la confidencialidad e integridad de los sitios web estatales.

Los centros encargados del control de ciberseguridad de los portales son el CERTuy y el SOC. El CERTuy se constituye de especialistas en técnicas de prevención y respuesta ante incidencias de seguridad en los sistemas informáticos, mientras que el segundo está integrado por analistas en seguridad cuyo fin es monitorear permanentemente los registros de actividad en la red interna de los institutos estatales.

2. Definición del problema.- El objetivo es optimizar la duración de análisis de logs de acceso en formato CLF correspondientes a sitios web por parte del equipo de respuesta (CERTuy) y operaciones en seguridad (SOC) de AGESIC frente a incidentes de ciberseguridad. Un incidente consiste un comportamiento inusual en un sitio web. Por ejemplo, una denegación de servicio (los usuarios no pueden acceder al sitio) o alteración del contenido del sitio web y operaciones no autorizadas a la base de datos del sitio.

Ante un incidente analistas del CERTuy piden al sector TI del ente afectado un conjunto de logs correspondientes a un mes. Seguidamente los analistas filtran los datos para identificar líneas anómalas que indican el momento en el que el atacante accedió al sitio en base a la fecha del incidente.

Considerando el diagrama basado en el trabajo de Skoudis [1] en la figura I, los analistas ya conocen el incidente una vez ocurrido. Son capaces de identificar el atacante en la etapa 5 (Alterar sitio) por los efectos producidos en un sitio web, pero les lleva más tiempo identificar el atacante en la etapa 2 y 3 (Escanear y obtener acceso). En otras palabras, los analistas están interesados en detectar más rápidamente ataques ocultos y persistentes en el tiempo. Estos ataques se llaman en la literatura como Advanced Persistent Threats (APT).

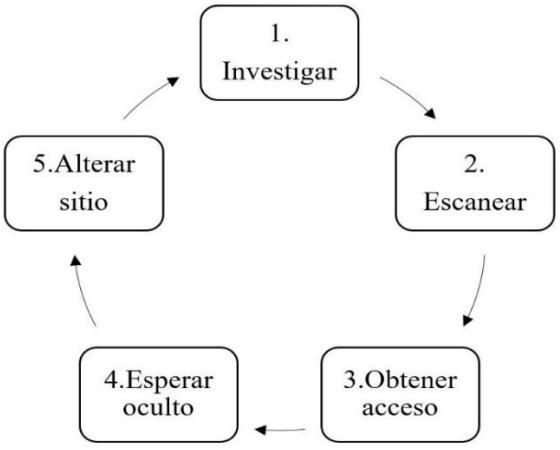

Figura I.- Etapas de un ataque según Skoudis (adaptación)

Los APT son costosos de detectar con antelación debido a que el comportamiento de los atacantes es similar al de usuario normales. Significa que los registros de actividad, llamados logs, de usuarios normales y atacantes no son inmediatamente distinguibles entre sí. Las

características del comportamiento de atacantes detectados hasta la fecha actual son variables y no siguen un patrón definido y acotado, sino que ellos durante meses recorren las páginas de los sitios web como usuarios normales, identificando los componentes gráficos y textuales del sitio web e intentan aprovechar vulnerabilidades de modo extendido en el tiempo. Por ejemplo, cierto día intentan ingresar instrucciones a través de un imagen y tres meses después usando otra dirección IP intentan una inyección SQL.

En resumen, ambos institutos requieren de un sistema que detecte los atacantes en las etapas de escaneo y obtención de acceso, con el fin de evitar con antelación que alteren los portales gubernamentales.

3. Objetivo.- Para resolver el problema planteado ambas instituciones solicitaron:

1. Investigar y brindar una solución automatizada que detecte los logs CLF que indican comportamiento anómalo. Además, la herramienta debe poder ser utilizable a logs CLF de cualquier sitio web.
2. Integrar la solución de 1 al sistema de gestión de eventos del SOC para afinar la precisión del sistema nivel web.

El requisito 1 se midió considerando la duración de filtrado de la solución, respecto el filtrado realizado manualmente por analistas, así como el porcentaje de líneas que indican comportamiento anómalo.

4. Antecedentes.- A la fecha del inicio del proyecto no se encontraron publicaciones para el filtrado y clasificación de registros de logs de servidores web en formato CLF, aunque si se han encontrado investigaciones para otros tipos de datos, por ejemplo, el cabezal HTTP, logs conexión TCP/UDP, o registros syslog.

A continuación, se mencionan publicaciones estudiadas para resolver problemas similares. Estas investigaciones aplican la misma metodología de resolución. Los logs se normalizan a un formato estándar y luego se convierten a vectores numéricos y a través de un modelo predictivo se clasifican los logs. En el caso del proyecto se probaron de modo experimental métodos de codificación para datos aproximados a los de CLF. Este paso previo de codificación es necesario porque los logs CLF contienen texto y caracteres no numéricos, y los algoritmos de aprendizaje automático requieren datos numéricos reales.

En la publicación de Zhang et al. [2] se intentó resolver un problema similar para detectar incidentes de ciberseguridad, en sitios web del Estado chino, estudiando el análisis del tráfico HTTP. Al igual que en este proyecto, Bertero et al. [3] buscó mejorar la rapidez de clasificación de logs. Sugirieron que el análisis manual de anomalías dura un tiempo excesivo. Para resolverlo propusieron técnicas de aprendizaje automático.

Tuor et al. [4] se basó en la detección de actividad anormal en logs de una aplicación con múltiples usuarios. De modo parecido a este proyecto, para clasificar adecuadamente el comportamiento de los usuarios, las líneas de actividad de cada usuario se convirtieron a un vector numérico. Por otra parte, Ming et al. [5] implementó un método de detección de ofensas a servidores web y Liu et al. [6] investigó detección de anomalías en aplicaciones compartidas de la Universidad Xi'an Jiaotong (China). Los datos utilizados fueron logs sintéticos.

5. Metodología de desarrollo de solución.- La solución propuesta es un clasificador de logs. Para obtener la solución especificada en la sección 3. se definieron tres etapas:

I. Implementación de un clasificador prototipo a medida.
II. Obtener un clasificador definitivo a partir de las correcciones del prototipo por parte de los analistas.
III. Integrar el clasificador al sistema de gestión de eventos de ciberseguridad usado por el SOC.

6. Solución implementada.- Con el objetivo de integrar el clasificador al sistema de gestión de eventos de ciberseguridad del SOC y debido a que se esperaban variaciones en especificaciones de la solución, además de que se valoraba que fuera una solución completamente controlable y confiable, se decidió desarrollar la solución desde cero. En la figura II se muestran los componentes y usuarios de la solución:

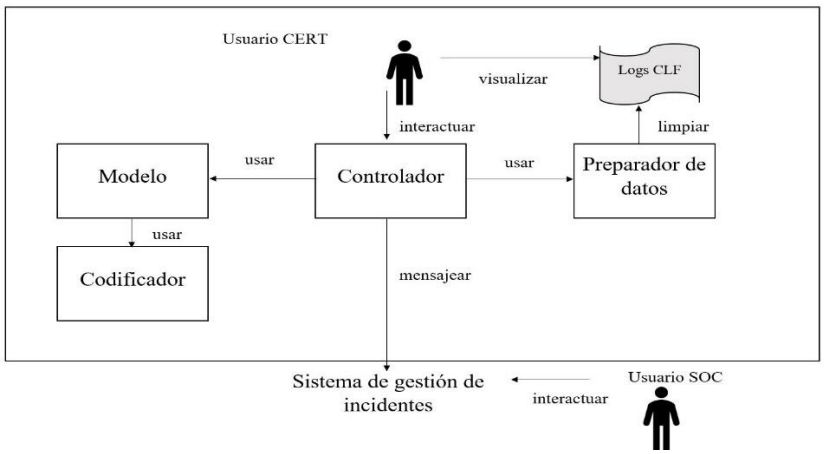

Figura II.- Diagrama de componentes y usuarios de la solución

7. Resultados.- Según lo mencionado en la sección 3, se establecen los umbrales para determinar la mejora en el proceso de respuesta por un analista experto. El clasificador logró los umbrales especificados y los superó. En la figura III se muestran los resultados.

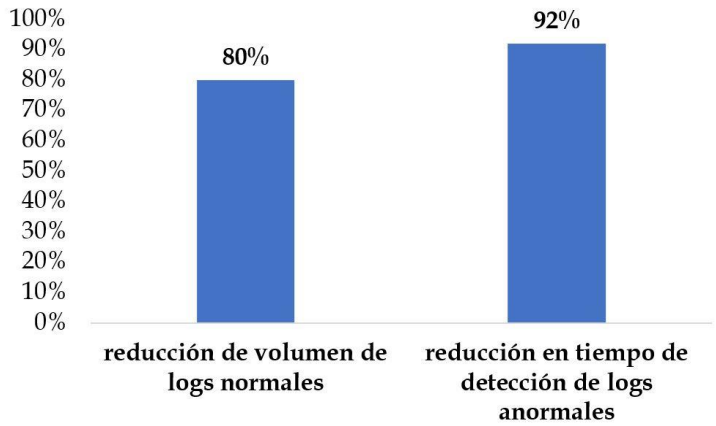

Figura III.- Resultados finales

7.1. Reducción de volumen de logs normales.- Debido a que se conoce la cantidad de falsos positivos, falsos negativos, verdaderos positivos y verdaderos negativos (los valores de la matriz de confusión) es factible medir de modo preciso esta métrica. En las pruebas el clasificador se ejecutó con 152 trazas, incluyendo ejemplos provistos por analistas del CERTuy. Para esa prueba se obtuvo un recall de 82%. Dada la matriz de confusión, 122 se clasificaron como verdaderas negativas (trazas clasificadas como normales que efectivamente contenían actividad usual). Así, se redujo 80% el volumen de logs normales respecto el total de datos de prueba. Se debe considerar que estos resultados aplican al conjunto de pruebas realizadas para los datos disponibles a la fecha de realizar la prueba. Para mantener estos resultados en el tiempo se debe reentrenar el modelo periódicamente.

7.2. Reducción en tiempo de detección de logs normales.- Este indicador se midió según la cantidad de horas requeridas para encontrar un volumen de 1 gigabyte de logs anormales. Según el criterio de los analistas se tardan 48 horas en detectar un conjunto de logs anormales de 7GB. Dado que el volumen de los logs CLF es de 28% respecto el total de datos registrados, contando otros formatos de datos, se asume que detectar 1,96GB de logs anormales CLF se tardan 13 horas. Por tanto, un analista tardaría 6,8 horas por gigabyte. El clasificador clasificó 450 MB en 15 minutos, por lo que tarda 0,56 horas (aproximadamente 30 minutos) en clasificar 1 gigabyte. Respecto la detección por un analista esto es una mejora del 92%. Sin embargo, se debe tener en cuenta que esta prueba no es extrapolable a volúmenes mayores de datos. En el caso de considerar un volumen mayor de datos, se debería realizar las pruebas adicionales pertinentes y en caso de ser necesario incorporar al clasificador herramientas para asegurar la escalabilidad.

8. Conclusiones.- El proyecto fue pionero, en el sector de ciberseguridad de AGESIC, en cuanto que se implementaron métodos y algoritmos de aprendizaje automático. Este trabajo fue un proyecto piloto para futuros proyectos que profundicen estas tecnologías. Los resultados obtenidos establecen una línea base que permita validar y medir el rendimiento y calidad de proyectos adicionales.

Para futuro quedan pendiente tareas de mantenimiento de la solución. Por ejemplo, establecer un equipo de trabajo encargado supervisar el entrenamiento periódico del modelo predictivo requerido. Adicionalmente se pueden investigar métodos más innovadores, tal como aquellos basados en aprendizaje no supervisado. En definitiva, existe potencial para extender y desarrollar lo introducido en este proyecto.

9. Referencias

[1] E. Skoudis and T. Liston, *Counter hack reloaded: a step-by-step guide to computer attacks and effective defenses*. Stoughton, Massachusetts, EEUU: Prentice Hall Press, 2005.

[2] S. Zhang, B. Li, J. Li, M. Zhang and Y. Chen, "A novel anomaly detection approach for mitigating web-based attacks against clouds," in IEEE 2nd International Conference on Cyber Security and Cloud Computing, New York, 2015, pp. 289-2942. [Online], Available: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7371496&isnumber=7371418

[3] C. Bertero, M. Roy, C. Sauvanaud, and G. Trédan, "Experience report: log mining using natural language processing and application to anomaly detection," in *28th International Symposium on Software Reliability Engineering*, Toulouse, 2017. [Online], Available: https://hal.laas.fr/hal-01576291/document

[4] A. Tuor, S. Kaplan, B. Hutchinson, N. Nichols, and S. Robinson, Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. Ithaca, Nueva York: Universidad de Cornell, 2017. [Online], Available: https://arxiv.org/abs/1710.00811

[5] Z. Ming, X. Boyi, B. Shuai, L. Shuaibing, L. Zhechao, L. Derong, X. Shengli, L. Yuanqing, and Z. Dongbin, "A deep learning method to detect web attacks using a specially designed CNN", *Neural Information Processing, International Conference on Neural Information Processing, Lecture Notes in Computer Science*, vol. 10638, pp. 828-836, 2017.

[6] Z. Liu, T. Qin, X. Guan, H. Jiang and C. Wang, "An integrated method for anomaly detection from massive system logs," *IEEE Access*, vol. 6, pp. 30602-30611, 2018.